MATHEMATICS AND THE MIND

Michael Redhead

Abstract

Granted that truth is valuable we must recognize that *certifiable* truth is hard to come by, for example in the natural and social sciences. This paper examines the case of mathematics. As a result of the work of Gödel and Tarski we know that truth does not equate with proof. This has been used by Lucas and Penrose to argue that human minds can do things which digital computers can't, viz. to know the truth of unprovable arithmetical statements. The argument is given a simple formulation in the context of sorites (Robinson) arithmetic, avoiding the complexities of formulating the Gödel sentence. The pros and cons of the argument are considered in relation to the conception of mathematical truth.

Mathematics and the Mind*

Michael Redhead

Centre for Philosophy of Natural and Social Science, LSE

Isn't it a bad thing to be deceived about the truth, and a good thing to know what the truth is? ------ Plato

---

Let me begin with a bold assertion. Ascertaining the truth in any field of enquiry is intrinsically good. Notice I am not denying that actions which employ or implement our knowledge may be morally bad. One only has to think of nuclear physics in the context of weapons of mass destruction. Nor am I denying that the truth about some impending calamity, for example, may lead to mental anguish which has to be balanced against the opportunity to display courage and fortitude Nor am I claiming that telling the truth to other people is always the right thing to do since one may have good reason to believe that they will employ the knowledge in a harmful way. Nor am I necessarily denying the fact/value distinction, that is a separate issue.

My claim is of course controversial: it advocates knowledge for its own sake. Many people might argue that *interesting* knowledge is good, not just knowledge per se. The truth about the exact number of hairs on my head might seem totally uninteresting but from the point of view of arguments about nature versus nurture might prove not to be so! So 'interesting' is something of a weasel word that is hard to pin down in any absolute sense.

But if you grant me that truth is a good thing, then reliable methods of ascertaining the truth will also be a good thing. In science, notoriously,

certifiable truth is hard to come by. In other fields, such as theology, claims to truth usually appeal ultimately to accepting some form of inspirational revelation, the authenticity of which can always be challenged, and in ethics, aesthetics, politics, economics and the social sciences generally truth claims seem equally problematic.

So let us turn to mathematics as a field in which the possibility at least of genuinely certifiable truth seems more promising.

At first glance truth in mathematics seems to equate with proving theorems. But we need to be careful here. There are basically two sorts of mathematics. An example of the first sort is group theory, where one is really concerned with 'unpacking' the logical consequences 'locked up' so to speak in the axioms of group theory. The truths of group theory are analytic in the sense that the conditional statement 'If the axioms are true then the theorems are true' is itself a logical truth. One is not asking the question: are the theorems true per se? Rather one is claiming, if the axioms are true under some interpretation, then the theorems are also true under that interpretation.

Contrast the situation in group theory with that in number theory. Here the hope is that the axioms are true of an *intended* interpretation, i.e. the natural numbers 0,1,2... Of course we know from the work of Gödel and Tarski, that it is not possible to prove or even, in a certain technical sense, to define, all the arithmetical truths. So truth outruns provability (or definability). But does this just mean that arithmetical truths outrun *knowability*? The curious thing is that this does not follow.

Gödel, for example, famously argued that there are true sentences of arithmetic (true that is in the intended interpretation) which cannot be proved

from the Peano axioms[1], and yet we can claim to know them to *be* true. This argument has been taken up by John Lucas and more recently by Roger Penrose to argue that minds can do things which computers cannot do, and hence that minds cannot be (digital) machines.[2] Of course we can always go to a stronger system of axioms to prove the Gödelian sentence, but then we can re-Gödelize, and so on, ad infinitum. Put succinctly,

> For any proposed Gödelian sentence there exists a machine which can deal with it (i.e. prove it) (1)

but

> It is not the case that there is a machine which can deal with all Gödelizations (2)

Essentially Lucas and Penrose are invoking (2) to resist the claim that minds are machines, whereas many people argue that (1) is all that is needed to establish exactly the reverse.[3]

But how do we know the Gödelian sentence G (which essentially announces its own unprovability in a suitable scheme of translation) really does express a truth about the natural numbers. There are various answers that can be given.

The simplest just says that if G is false it would be provable and hence true, whence by reductio it cannot be false. But this presupposes that the axioms of Peano arithmetic are true. So how do we go about convincing our sceptical opponent that that is the case.

Or again, it is often argued that one can prove in first-order Peano arithmetic the conditional statement

$$CON \rightarrow G \qquad (3)$$

where 'CON' expresses the consistency of the Peano axioms. Now we cannot prove CON, since if we could we could prove G, which by Gödel's first incompleteness theorem we cannot do (this is just the second incompleteness theorem). But if Peano arithmetic is sound then (3) is true, and we can argue *inductively* for the truth of CON (no-one has so far discovered a contradiction in Peano arithmetic). But combining this inductive argument for the truth of CON with the truth of (3), we can now detach the consequent in (3) and infer the truth of G. But this argument again presupposes the soundness of Peano arithmetic and in particular the truth of its axioms.

I want now to consider this problem of how we know the truth of the axioms and in so doing show how the Lucas-Penrose argument can be recast in a much simpler and more transparent form.

The Peano axioms are of three sorts. First there are axioms of the form:

1. Zero is a number.
2. Every number has a unique successor.
3. If two numbers have equal successors then the numbers are equal.
4. Every number other than zero is the successor of some number.

These axioms are arguably analytic in the sense that they express defining properties of the numbers 0, 1, 2, 3 ... . If any of these axioms were false we would not be talking about numbers.

But then there is the notorious fifth axiom, the induction axiom (more accurately an axiom schema in first-order logic).

This says that for any admissible predicate[4] F, if F is true of zero and if, given that F is true of n then it is true of the successor of n, then it is true for all n.

Poincaré (1952) famously claimed that this was not an analytic truth, but forced itself on us with such conviction, that it was a candidate for the elusive synthetic a priori.

The induction axiom certainly cannot be proved from the first four axioms, so what is the source of the conviction that Poincaré talks of? We shall return to this in a moment.

The remaining axioms of Peano arithmetic introduce recursive definitions of addition and multiplication, essentially defining addition in terms of repeated application of the successor operation, and multiplication as repeated addition. These axioms can also be arguably regarded as analytic.

To appreciate the significance of the more mysterious fifth axiom, it is useful to consider a weaker form of arithmetic that Lucas calls sorites arithmetic which employs all the axioms of Peano *except* the induction axiom. A carefully formulated version of sorites arithmetic is called (following Raphael Robinson) the system Q in the standard text-book of Boolos and Jeffreys (1980). Sorites arithmetic essentially allows us to move from one number to the next and repeat the operation a finite number of times. So we can formulate the usual procedures of, for example, long multiplication and verify statements like 3 x 6 = 6 x 3 or 279 x 631 = 631 x 279. Indeed we can check

out the commutative property of multiplication for any pair of numbers m and n. But the proof gets longer and longer as m and n get bigger and bigger.

So, in sorites arithmetic

For all pairs (m,n) there exists a proof that

$m \times n = n \times m$ (4)

But what we cannot do is to switch the order of the universal and existential quantifiers to get

There exists a proof that for all pairs (m,n)

$m \times n = n \times m$ (5)

(5) expresses the proof of the commutative law of multiplication in arithmetic.

But for (5) we need the induction axiom. Why is this? Because the proof referred to in (4) depends on the pair of numbers chosen, and while always remaining finite in length, increases in length without limit as the numbers get bigger and bigger. So there is no *finite* proof that will work for *all* pairs of numbers, which is what we need for (5).

But we can argue that

For all pairs (m,n), $m \times n = n \times m$ (6)

is nevertheless true in sorites arithmetic.

We have argued that the axioms and hence the theorems of sorites arithmetic are analytically true, so we can replace (4) by

For all pairs (m,n) it is true that $m \times n = n \times m$ (7)

But (7) is strictly equivalent to

It is true that for all pairs (m,n), m x n = n x m (8)

But (8) just says that the commutative law of multiplication is true.

Effectively truth commutes with the universal quantifier, whereas provability doesn't!

So we have here a case in which certifiable truth outruns provability.[5]

The same argument applies to many other theorems that can't be *proved* in sorites arithmetic, such as the commutative law of addition or the associative laws of addition and multiplication, and, more particularly it applies to establishing the truth as against the provability of the induction axiom itself.[6]

So here is a simple argument that human minds can know the truth of statements which can be expressed or represented in a system but can't be proved in the system, but without employing the complexities of understanding the proof of Gödel's theorem!

Of course we can go on to demonstrate that every consistent, axiomatizable, extension of sorites arithmetic is incomplete in the sense that there are true sentences of arithmetic which cannot be proved in *any* such extension. This requires the much more complex Gödelian arguments.[7]

We have only taken the first step of showing that sorites arithmetic itself is incomplete. But the important ingredient is to see how truth *can* outrun provability in the interesting sense that this is not the case just for analytic truth, but for truths like the induction axiom, which are in no clear sense constitutive of the *meaning* of number [8].

Rather our argument depends on having a concept of truth. In particular it works for any conception of truth which conforms to basic Tarskian semantics.

But can I now conclude that minds are not machines?

If you are moved by constructivist considerations, you may want to deny that the truth of statements involving universal quantification over infinite domains has any clear meaning. But, for a Platonist, and that means most 'working' mathematicians, the argument may seem compelling.

But now we have come full circle. If you accept you are not a machine, then this at the very least opens the door to the possibility of how you can *understand* the evaluative claim I started with (even if you don't agree with it), that ascertaining the truth is intrinsically good, something that no machine presumably could do!

Acknowledgements

I am grateful to John Lucas, Richard Sorabji , Arnie Koslow and Richard Healey for helpful discussions.

Notes

* Paper contributed to the Conference entitled The Place of Value in a World of Facts, to be held at the LSE in October 2003.

1. Gödel assumed the Peano axioms to be ω-consistent. This was later relaxed by Rosser to the weaker assumption of consistency. See Machover (1996).

2. See Lucas (1961) and (1970) and Penrose (1989) and (1994). Penrose employs a simplified form of the Gödel argument due to Turing.

3. For complete references to the many different types of response to the Lucas-Penrose argument see Lucas (2000), Penrose (1994).

4. What predicates are admissible depends on whether we choose to employ first or second-order logic.

5 In more formal terms, we have sought to demonstrate that sorites arithmetic is ω-complete in a semantic sense, while recognizing that it is ω-incomplete in syntactic terms.

From this perspective G$\ddot{o}$del's theorem establishes the ϖ–incompleteness of Peano arithmetic. The truth of the G$\ddot{o}$del sentence then follows from an argument similar to the one given in the text showing that Peano arithmetic is ϖ –complete in a semantic sense. Trivially, the fact that truth outruns provability follows from bivalence for any formal system which is incomplete. But of course we are interested in knowing which of 'p'and 'not p' is true.

6 It is important to notice that none of these theorems can be proved using the logical principle of Universal Generalization (UG). UG allows us to

pass from the fact that a *typical* member of a collection possesses a property to the claim that all members of the collection possess that property. But the decisive point is that, for the purposes of these theorems, all the numbers are uniquely different – there is no such thing as a typical number.

We stress again that we are talking of theorems which are true of the intended interpretation. There will, of course, be non-standard interpretations of sorites arithmetic, but these will include an initial segment isomorphic to the natural numbers. It is with regard to this initial segment that we are claiming the theorems to be true. I owe this point to Richard Healey.

An additional remark: these theorems could be proved if we employed an infinitary logic incorporating the so-called ϖ -rule. But such logics cannot of course be implemented on a machine.

7. See, for example, Boolos and Jeffrey (1980).

8. Sorites arithmetic is the weakest system that allows this possibility, i.e. for still weaker systems truth outruns provability only for analytic statements.

References

Boolos, G.S. and Jeffrey, R.C. (1980): *Computability and Logic*, 2nd ed, Cambridge: Cambridge University Press

Lucas, J.R. (1961): 'Minds, Machines and Gödel', *Philosophy*, *36*, 112-127

Lucas, J.R. (1970): *The Freedom of the Will*, Oxford: Oxford University Press

Lucas, J.R. (2000: *The Conceptual Roots of Mathematics*, London: Routledge

Machover, M. (1996): *Set Theory, Logic and their Limitations*, Cambridge: Cambridge University Press

Penrose, R. (1989): *The Emperor's New Mind*, Oxford: Oxford University Press

Penrose, R. (1994): *The Shadows of the Mind*, Oxford: Oxford University Press

Poincaré, H. (1952): 'On the Nature of Mathematical Reasoning', in *Science and Hypothesis*, New York: Dover, pp.1-19. Reprinted in P. Benacerraf and H. Putnam (eds.), *The Philosophy of Mathematics*, 2nd ed, Cambridge: Cambridge University Press, 1983, pp.394-402